MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11-13, 2002, Nara-ken New Public Hall, Nara, Japan

13—16

Facial Parts Recognition by Hierarchical Tracking from Motion Image and Its Application

Tsuyoshi Yamaguchi[1] SCCS Chukyo University
Takuma Funahashi[2] SCCS Chukyo University
Masafumi Tominaga[3] HOIP SOFTOPIA JAPAN
Hiroyasu Koshimizu[4] SCCS Chukyo University

Abstract

We have been developing the methods for facial parts recognition such as eyes, a nose, a mouth, etc. This paper pays attention especially to eye (iris) and nostril recognition by Hough transform, and new hierarchical method for face tracking and recognition is proposed in order to realize a human friendly facial interface media eye-contacting to the partner on TV conference environment.

1 Introduction

We have been developing the methods for facial parts recognition such as eyes, a nose, a mouth, etc. This paper pays attention especially to eye (iris) recognition by Hough transform, and new hierarchical method for face tracking and recognition is proposed in order to realize a human friendly facial interface media eye-contacting to the partner on TV conference environment.

Hierarchical tracking consists of facial tracking from motion images, iris tracking from the tracked frame, facial parts region tracking guided by the irises and nostrils, and facial parts contour tracking in the respective facial parts region. Facial tracking is performed by using a PTZ (Pan-Tilt-Zoom) camera together with the fixed CCD camera. PTZ camera can capture a series of facial images with a sufficient spatial resolution by being controlled by the position and size of the face region extracted in CCD camera images. In this way, this hierarchical system gives a human friendly free interface space to the testee in front of the camera. And by analyzing the series of facial images from PTZ camera, a pair of the irises and a pair of nostrils are extracted in more than 96% and 92% success rates, respectively. Based on the information of the irises and nostrils, rectangular regions in which the facial parts are contained are tacked. Finally the contour of the facial parts is extracted from the rectangular region. Since the relationship among TV camera, monitor display and a human in front of the camera is easy to model geometrically by the extracted successive facial images, we proposed a KANSEI facial images which are eye-contacting to the partner on the TV monitor in the TV conference for the better human interface media on the network environment.

2 Hierarchical Tracking

2.1 Principle

We have been developing the methods for facial parts recognition such as eyes, a nose, a mouth, etc. This paper pays attention especially to eye (iris) recognition by Hough transform, and new hierarchical method for face tracking and recognition is proposed in order to realize a human friendly facial interface media. Fig.1 is a conceptual figure of the hierarchical tracking system.

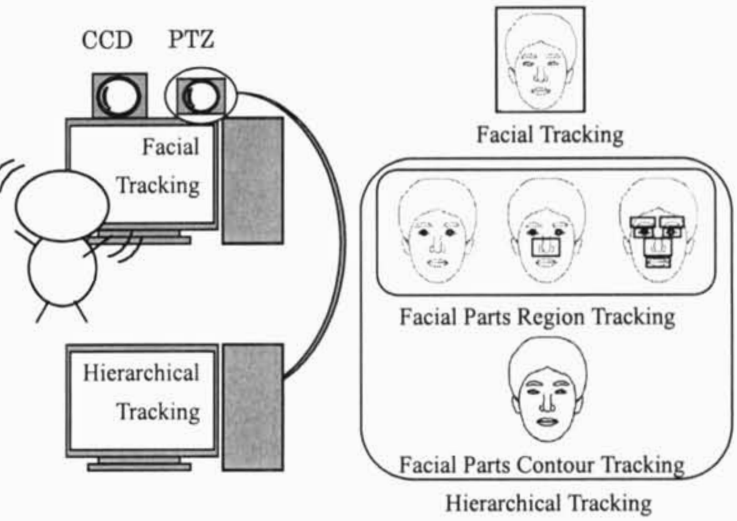


Fig.1 facial tracking

2.2 Facial Tracking

Facial tracking is performed by using a PTZ camera together with the fixed CCD camera. As shown in Fig.1, PTZ camera can capture a series of facial images with a sufficient spatial resolution controlled by the position and size of the facial region extracted in CCD camera images.

First, facial region tracking was realized by the template matching for input frame. The normalized correlation given by eq.(1) was utilized in the template matching.

$$R(m,n)=\frac{\sum_j\sum_i (T_{ij}-Tf_{ij})(F_{i+m,j+n}-\overline{F})}{\sqrt{\sum\sum[T_{ij}-\overline{T}]^2 \bullet \sum\sum[F_{i+m,j+n}-\overline{F}]^2}} \quad (1)$$

$$\left(\overline{T}=\frac{\sum\sum T_{ij}}{N} \quad \overline{F}=\frac{\sum\sum F_{i+m,j+n}}{N}\right)$$

In order to cope with the changes of facial expression and facial size, the template is always updated by a way of frame to frame. This will provide a movable space to a person in front of the camera. The new template for the next frame is generated by the following algorithm. The

1 Address: SCCS, Chukyo University, 101 Tokodachi, Kaizu-Cho, Toyota, 470-0393 Japan. E-mail: **tsuyoshi@koshi-lab.sccs.chukyo-u.ac.jp**
2 Address: SCCS, Chukyo University, 101 Tokodachi, Kaizu-Cho, Toyota, 470-0393 Japan. E-mail: **takuma@koshi-lab.sccs.chukyo-u.ac.jp**
3 Address: HOIP, Softopia Japan, 4-1-7 Kagano, Ogaki City, Gifu 503-8365 Japan. E-mail: **tomy@hoip.jp**
4 Address: SCCS, Chukyo University, 101 Tokodachi, Kaizu-Cho, Toyota, 470-0393 Japan. E-mail: **hiroyasu@sccs.chukyo-u.ac.jp**

current frame is subtracted by the background image and binalized so that the horizontal and vertical sizes of the template can be decided by the XY projections. Next, the vertical and horizontal amount of PTZ camera movements is decided on the basis of the center coordinates of the extracted facial region template. The zooming parameter Z of PTZ camera is decided by eq.(2) with the template size L.

$$Z = \frac{Z\ \mathrm{max}}{\mathrm{Lmax}\ -\mathrm{Lmin}} \times (\mathrm{Lmax}\ - \mathrm{L}) \qquad (2)$$

The facial tracking by PTZ camera is performed by transmitting the obtained information to the PTZ camera. Fig.2 (a) is a series of the results of facial region extracting. Fig.2 (b) is the same series of the results of facial tracking by PTZ camera.

(a) facial region extracting (b) facial tracking

Fig.2 facial tracking by PTZ camera

2.3 Facial Parts Region Tracking

2.3.1 Principle

Facial parts region tracking is leaded by the iris recognition. Iris recognition was composed by Hough transform for circle detection applied to the binary image provided from PTZ camera. Basing on a pair of irises, a pair of nostrils is extracted in the same way. Being controlled by these two pairs of circles, facial parts regions for eyebrow, eye, nose, mouth, can be automatically defined adaptively to the successive images. Fine contours of the facial parts could be finally extracted by investigating exclusively in these regions.

2.3.2 Iris Recognition

A method for the recognition of irises from the gray images is proposed by using Hough transform for circle detection. Several candidates of a pair of the irises were extracted at first by applying Hough transform for circle detection to the binary image. The binarization method was especially constructed by [1], [4]. The voting ranges of the parameter space (a,b,r) were limited to some extent in order to reduce the computation cost and to enforce the performance. Parameters a and b indicate the center of iris. Parameter r indicates the radius of the iris. The best pair of the irises is detected from the candidates in accordance with the criteria standards given by that the number of votes is bigger, that the positional relation between left and right irises is horizontal and that the radius of the left equals to the right. Therefore the recognition procedure was designed as follows:

Step1. Set K as the threshold for the peak detection.
Step2. Detect the coordinates of irises whose peak is greater than or equal to K.
Step3. Prepare the list of all candidates of the irises.
Step4. Choose the candidates pairs whose vertical distance is smaller than the threshold.
Step5. Choose the candidates pairs from the list whose radius are equal.
Step6. Among the rest of the list, extract a pair of right and left irises whose horizontal distance is minimum.

In experiment, the facial database made by Softopia Japan was used in order to check the basic performance of this algorithm. This facial database consists of 300 images. (the age section of 5 years old, and the number of it is 15 men and women in each section.)

It is important to evaluate the extraction accuracy by using the well-regulated database with various attributes from the meaning of the mathematical statistics. Fig.3 shows an example of the iris recognition result. As given in table 1, the iris recognition was successful in 291 images among 300 images. Although more than 96% success rate of iris recognition was provided for the frontal face shown in Fig.4, it is hard to guarantee the performance for the various aspects of the face. There are not only frontal faces but facial images taken by changing the aspects of 15 degrees of four directions in the above-mentioned facial database. Then, we used these facial images to estimate the iris recognition accuracy over the head pose inclination. Fig.4 shows an example of the iris tracking result. Even if the aspect of the head is leaned in 15 degrees, it was clarified that the tracking could be carried out at the high success rate of the extraction.

Table 1 success rate of iris recognition

	Male	Male (Glasses)	Female	Female (Glasses)	Total
○	104/108 96.2%	38/42 90.4%	137/137 100.0%	12/13 92.2%	291/300 96.9%
△	0/108 0.0%	1/42 2.3%	0/137 0.0%	0/13 0.0%	1/300 0.3%
×	4/108 3.7%	3/42 7.1%	0/137 0.0%	1/13 7.6%	8/300 2.7%

Fig.3 iris recognition (result)

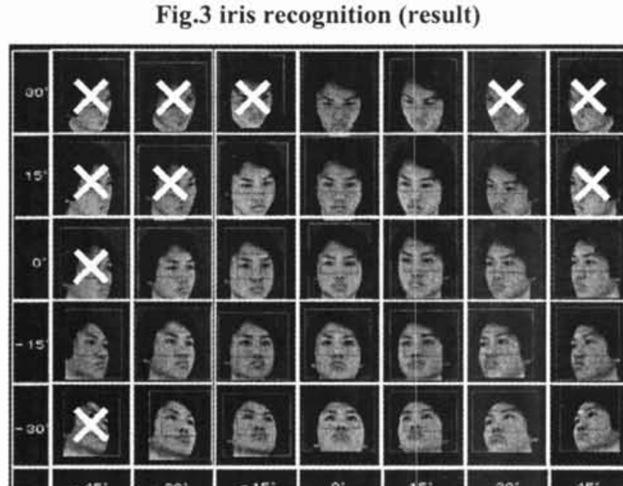


Fig.4 iris tracking result for the leaned faces

2.3.3 Nostril Recognition

Next, the nostrils were extracted by the same method of iris recognition in nose region decided on the basic of the position of a pair of irises.
The nose region shown in Fig.5 was decided by eq.(3).

$$(nose_1.x,\ nose_1.y) = (left\_iris.x,\ left\_iris.y + \alpha)$$
$$(nose_2.x,\ nose_2.y) = (right\_iris.x,\ nose1.y + \beta) \qquad (3)$$

Fig.6 shows a couple of examples of the nostril recognition. The recognition rate was 92.3% (277/300) in the 300 images of the same database.

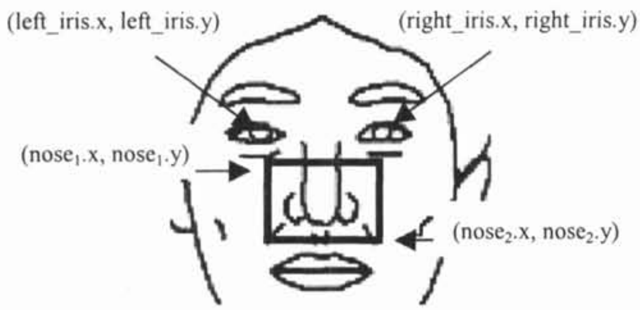


Fig.5 Nostril recognition

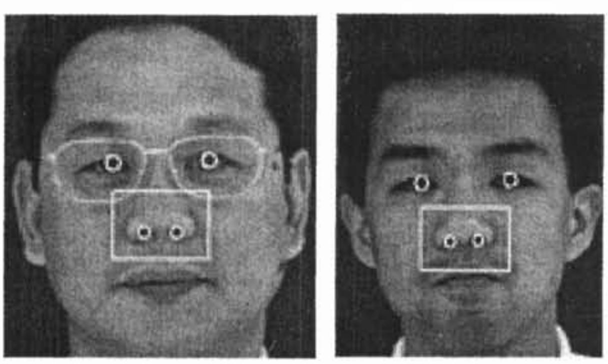


Fig.6 Nostril recognition (result)

2.3.4 Facial Parts Region Tracking

Next, the facial parts region is decided on the basis of the position of a pair of irises and of nostrils.
Fig.7 shows an example of facial region tracking result.

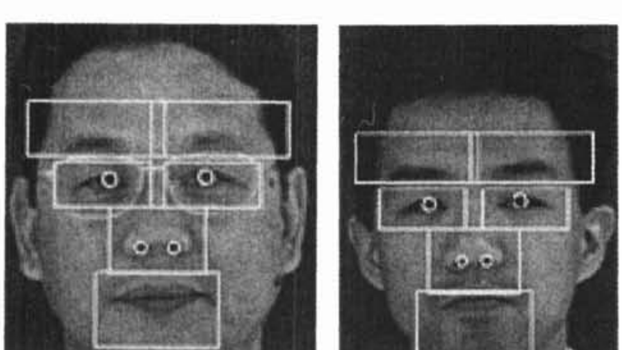


Fig.7 facial parts region tracking result

2.4 Facial Parts Contour Tracking

Next, facial parts contour tracking is performed by observing the facial parts region obtained in the section 2.3. The initial set of feature points is specified according to manual input to the first frame of the motion images. It is realized by the frame difference on the basis of the position of this feature point. The feature point tracking is performed based on the facts of that this difference pattern exists in the neighboring domain of each feature point and of that the nearest feature point should be extracted as the moved point. The contour tracking of the facial parts is performed by repeating this processing. Fig.8 shows an example of the series of the contour tracking.

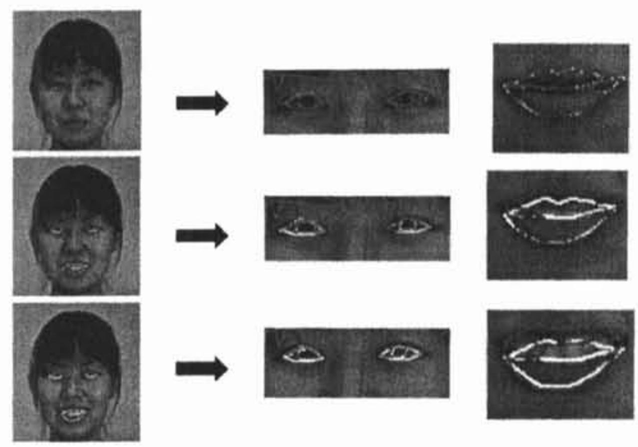


Fig.8 facial parts contour tracking result

3 Availability of Hierarchical Tracking

3.1 Principle

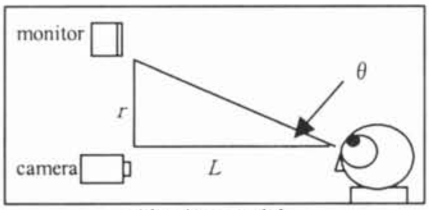
The new hierarchical method for face tracking and recognition is proposed in order to realize a human friendly facial interface media which is eye-contacting to the partner on TV conference environment.

3.2 Eye-contact Camera system

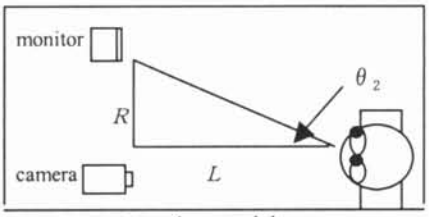
In order to model the situation of the TV conference environment shown in Fig.9, the parameters R and r are specified for modeling the vertical relation between the TV camera and the monitor, and the parameter L is specified for the horizontal relation between them. In the beginning, let us imagine the iris moves from the coordinate (x_0, y_0) extracted before to the new coordinate (x_1, y_1). This new coordinate (x_1, y_1) can be easily calculated by eq.(4) and eq.(5) characterized with the parameters θ and θ_2 indicating the spatial relationship among a person, camera and monitor. In this expression, functions Δx and Δy are designed to convert the parameter θ to the number of the pixels in the facial image.

$$\theta = \tan^{-1}\frac{R}{L} \qquad \theta_2 = \tan^{-1}\frac{r}{L} \qquad (4)$$

$$x_1 = x_0 + \Delta x(\theta) \ :\ \Delta x(\theta) = \theta/10$$
$$y_1 = y_0 + \Delta y(\theta_2) \ :\ \Delta y(\theta_2) = \theta_2/10 \qquad (5)$$



side view model



top view model

Fig. 9 parameters of eye-contact camera system

The center coordinate (x_1, y_1) and the radius r_1 of the extracted iris are utilized to regenerate the moved iris. Beforehand the eye regions are recognized by utilizing the extracted eye center coordinate (x_0, y_0). The eye regions (the white of eye, iris, skin and contour of the eyelid) are extracted for both of eyes, and basing on the recognition results, the moved iris is generated within the contour of the eyelid.

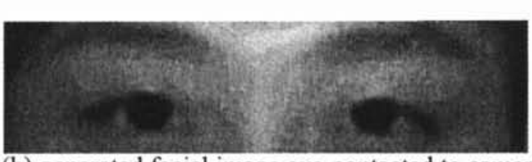
The pixels (x,y) in the region for new iris are painted in black at the region where the distance d between (x_1, y_1) and (x,y) is less than or equal to the radius d_0. All pixels (x,y) within the contour of the eyelid are painted in white at the region where the distance is greater than or equal to r_1. The black and white colors are decided as follows:

$$\begin{aligned} black &= \min \{F_{ij} \mid f_{ij}=1\} \\ white &= \max \{F_{ij} \mid f_{ij}=0\} \end{aligned} \quad (6)$$

After this procedure, the final irises are regenerated by smoothing processing.

We applied this system to the input facial image shown in Fig.10(a), and then new facial image given in Fig.10(b) was provided in such a way that the directions of the irises are slightly moved upward. Even though the details in the regenerated eyes given in Fig.10(b) seem not perfect, we can look naturally at the regenerated facial image shown in Fig.10(b) especially in the movie demonstration.[12] We performed a questionnaire experiment for evaluating the naturality of the generated movie, and we got successful reports on it.

(a) real facial image



(b) generated facial image eye-contacted to camera

Fig.10 eye-contact camera system result

4 Conclusions

This paper proposed new hierarchical facial parts tracking. Especially in this paper, we have presented the method for face tracking and recognition, and we have shown the possibility for realizing human friendly TV conference system. Future problems are to improve the regeneration method of irises, to introduce the real time processing method (currently 7 fps by Pentium Ⅲ 800MHz), and to cope with the motion of the head.

5 Acknowledgement

Authors wish to express thanks to IMS International HUTOP Research Project, Daiko Foundation Academic Research Promotion, Grant in Aid for Scientific Research Promotion of The Ministry of Education and HRC Research Promotion for supporting a part of this research.

References

[1] Oshida, M., Tominaga, M., Endoh, T., Toryu, T., Murakami, K., and Koshimizu, H. : "Facial parts recognition by Hough Transform", 1998 Annual meeting of Tokai-branch of IES, IEICE, D-708, P.354 (1998.9)

[2] Yamaguchi, T., Tominaga, M., Murakami, K., and Koshimizu, H. : "Generating Facial Images Eye-Contacting with Partner on the TV Conference Environment", Japan Academy of Facial Studies, 1999 Forum Faces, B-4P, P.25 (1999.8)

[3] Yamaguchi, T., Tominaga, M., Murakami, K., and Koshimizu, H. : "Generating Facial Images Eye-Contacting with Partner on the TV Conference Environment", 1999 Annual meeting of Tokai-branch of IES, IEICE, G-745, P.373 (1999.9)

[4] Segawa, Y., Endoh, T., Toryu, T., Murakami, K., and Koshimizu, H. : "Contour Extraction of face and irises for facial caricaturing ", Trans Image Information Media Society, Vol.51, No.11, pp.1935-1944 (1997.11)

[5] Ueno, T., Andoh, S., and Nakajima, M. : "A Method of Making Front Face for Eye Contact in Teleconference System", The 5th Symposium on Sensing via Image Information, SII'99, pp.71-76 (1999.6).

[6] Osamu Hasegawa, Shigeo Morishima and Masahide Kaneko : Face Information Processing, Trans.IEICE , Vol.J80-A, No.8, pp.1231-1249 (1997)

[7] Patrick Henry Winston : The Psychology of Computer Vision, McGraw-Hill computer science series (1975)

[8] Nagashima, Y. : Present State and Issues of SiLE ("Almost all interface media is face even in the hand gesture conversation."), Invited talk, PRMU99 – 141 (Nov.19, 1999)(Okinawa)

[9]www.brl.ntt.co.jp/info/kameg/emr-nc/vesearches.html

[10] Tominaga, M., Fujiwara, T., Murakami, K., and Koshimizu, H. : "On KANSEI Facial Image Processing For Computerized Facial Caricaturing System PICASSO", Proc. SMC'99 (Oct. 1999) (Tokyo)

[11] Funahashi, T., Uchi, A., Watanabe, T., Matsumoto, S., Tominaga, M. and Koshimizu, H. : "Face Tracking System with Fixed and PTZ Cameras" , Proc. 2nd Workshop for Practical Applications of Motion Image Processing, (Mar. 8-9, 2001) (Keio Univ., Yokohama)

[12] www.koshi-lab.sccs.chukyo-u.ac.jp/~tsuyoshi/reseach/